

Data Intake Report

Name: G2M insight for Cab Investment firm

Report date: 14/01/2023

Internship Batch: LISUM17

Version:

Data intake by: Rayhanul Islam Rumel

Data intake reviewer:

Data storage location: <https://github.com/rirumel/G2M-insight-for-Cab-Investment-firm>

Tabular data details: Cab_Data.csv

Total number of observations	359392
Total number of files	1
Total number of features	7
Base format of the file	.csv
Size of the data	20.1 MB

Tabular data details: City.csv

Total number of observations	20
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	759 bytes

Tabular data details: Customer_ID.csv

Total number of observations	49171
Total number of files	1
Total number of features	4
Base format of the file	.csv
Size of the data	1.00 MB

Tabular data details: Transaction_ID.csv

Total number of observations	440098
Total number of files	1
Total number of features	3
Base format of the file	.csv
Size of the data	8.58 MB

Proposed Approach:

- There were no null or duplicate values in the four datasets. Analyzed each dataset as well as merged and Mastered dataset using different analyzing and visualizing techniques in python. The notebook is well documented.
- In the instruction, the assignment asked us to consider the date range starting from 31-01-2016 to 31-12-2018. But, if I map the data using the instruction, I was getting data from the first 29 days of January 2019 as well. Therefore, I decided to consider 01-01-2016 instead of 31-01-2016 and it fitted well within the range.